

Sommaire

- ◆ Définition d'un entrepôt de données de la recherche
- ◆ Etude des solutions logicielles existantes
- ◆ Planning prévisionnel

Un entrepôt de données de la recherche...

... sert à stocker, décrire, référencer, préserver et exposer les données de la recherche produites par les chercheurs et les institutions de recherche.

Plan national science ouverte 2018 : demande de FAIRisation des données au niveau national.

<https://www.ouvrirlascience.fr/plan-national-pour-la-science-ouverte/>

Politique de science ouverte de l'Unistra : création d'un entrepôt de données

<https://scienceouverte.unistra.fr/strategie/pilier-2-fair-open-data/>

Un entrepôt de données de la recherche...

C'est un outil...

- De citation des données (DOI)
- De partage de données
- De visibilité pour les travaux de recherche
- Qui assure une qualité des données (FAIR, curation et certification)

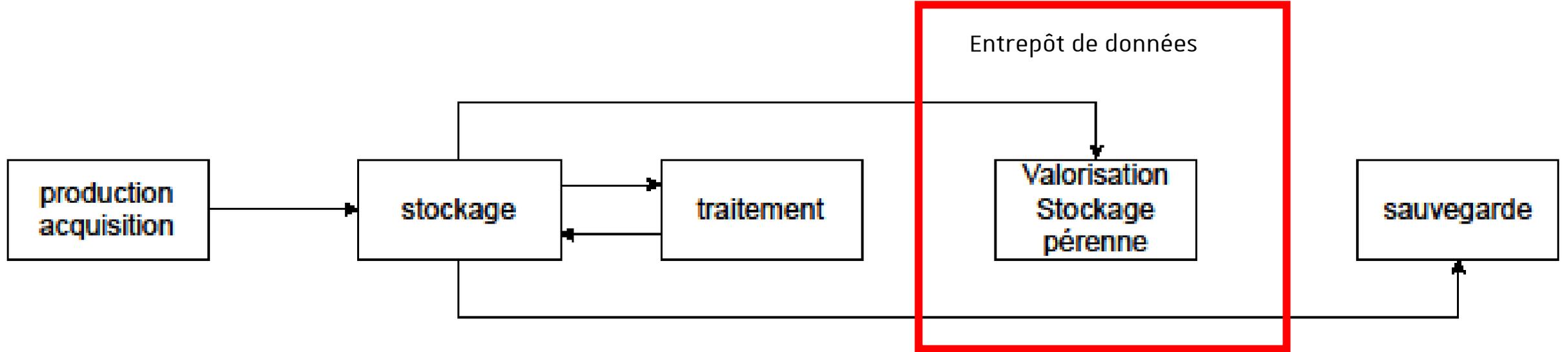
Ce n'est pas...

- ◆ Un FTP/NAS
- ◆ Un stockage privé

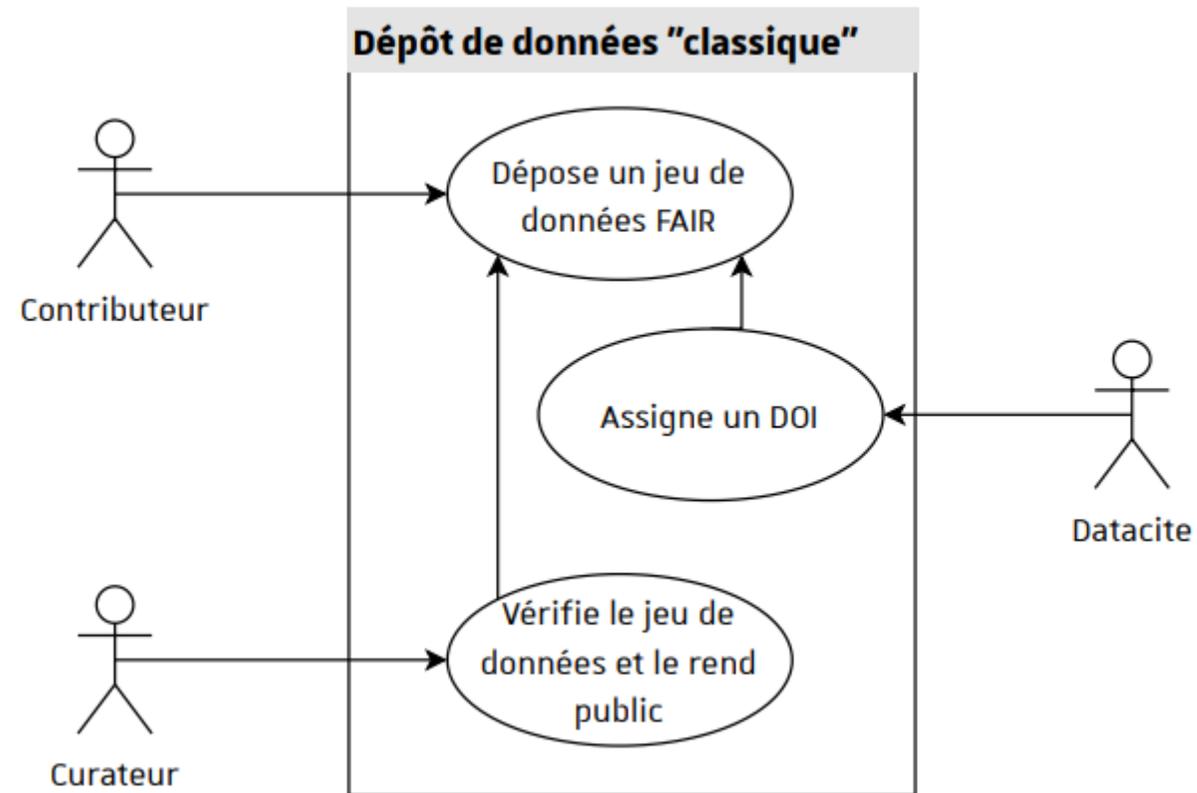
Ce qui existe

- ◆ Institutionnels, thématiques, disciplinaires
- ◆ Local (hébergé sur les serveurs de l'institution) ou à distance (sur des serveurs distants, dans d'autres régions, pays)
- ◆ commercial ou open source

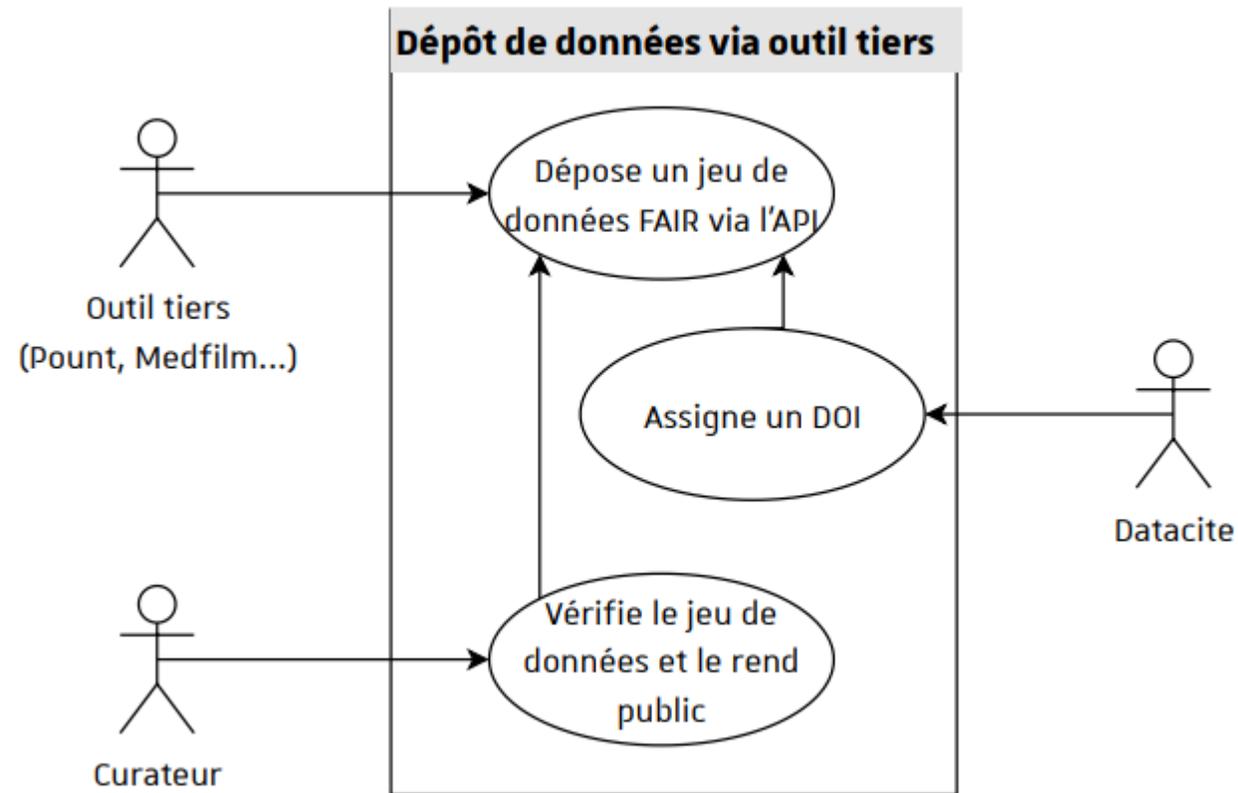
Workflow des données de la recherche



Dépôt d'un jeu de données



Dépôt d'un jeu de données



Les critères :

- ◆ Entrepôt multidisciplinaire
- ◆ Interopérabilité obligatoire avec d'autres types de solutions (via API)
- ◆ Compatible avec le socle technique de la Dnum (Python, s3....)
- ◆ Eviter que les données se retrouvent sur un cloud
- ◆ Equipe de développement fiable
- ◆ Communauté forte en cas d'open source

Caractéristiques de l'étude

- ◆ Etude menée par SBU/DNUM d'avril à septembre 2020, avec points d'étape avec Yannick Hoarau, porteur politique du projet, et échanges réguliers avec l'Université de Lorraine (projet semblable)

- ◆ 5 solutions étudiées
 - 1 solution commerciale
 - 4 solutions open source

- ◆ Effectuée sur 52 critères/fonctionnalités
 - critères métiers + techniques
 - tableau récapitulatif disponible sur demande

Entrepôts candidats

Commercial :

- ◆ Figshare (Digital Science)

Open source :

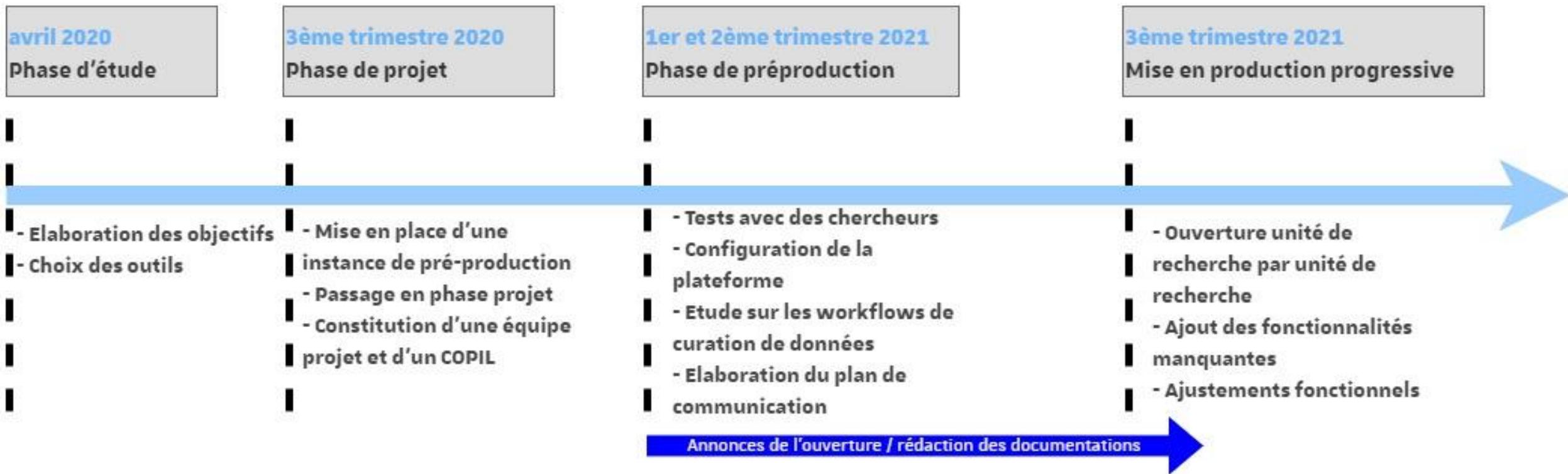
- ◆ Dryad (multi-universités, Ruby on Rails)
- ◆ Dspace CRIS (dSPACE GmbH et 4science, Java)
- ◆ InvenioRDM (CERN, Python)
- ◆ Dataverse (Harvard, Python)

Conclusion

Les développements de InvenioRDM et Dryad sont intéressants et il faudra certainement étudier ces solutions dans les prochaines années.

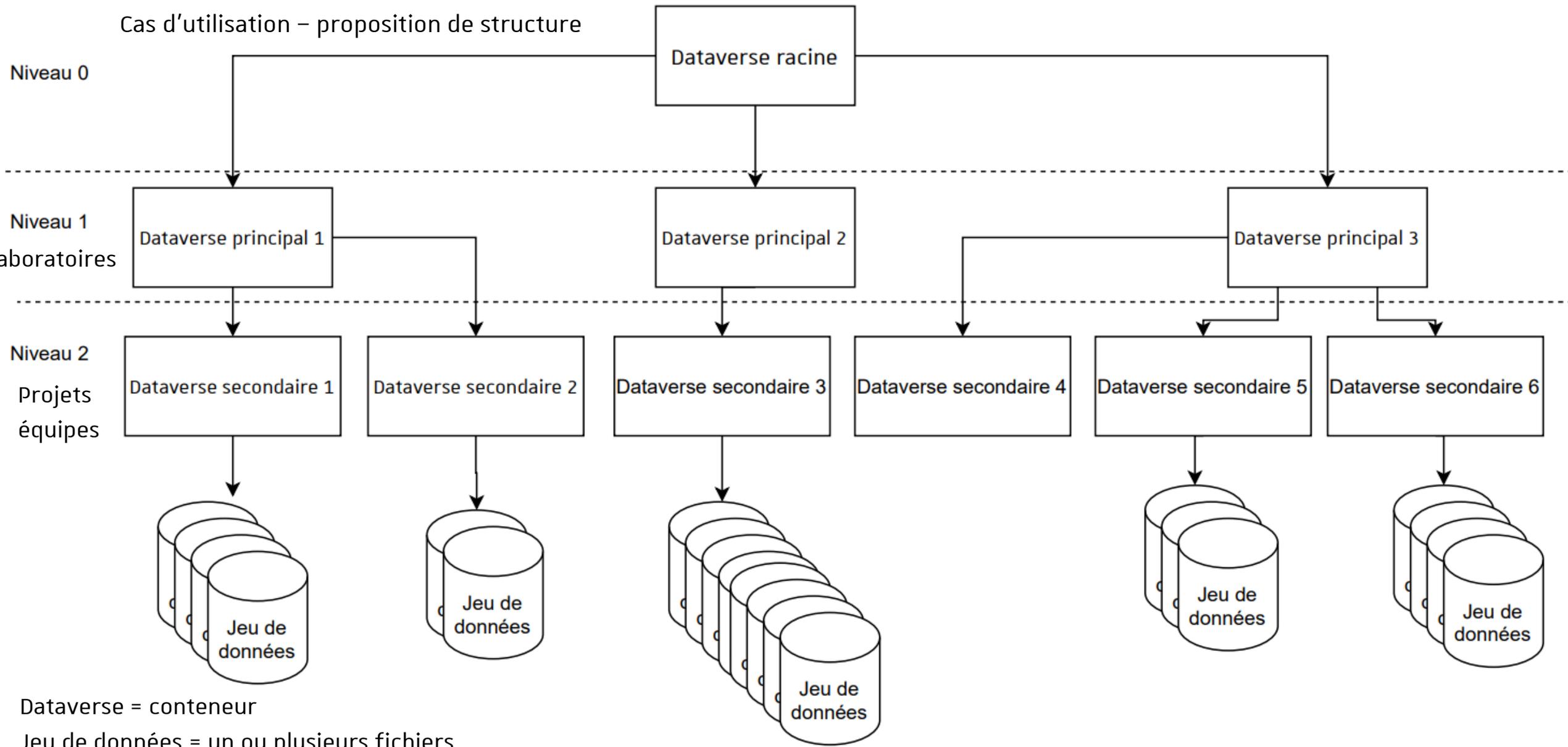
Mais Dataverse est le seul candidat valide à l'heure actuelle.

Planning prévisionnel entrepôt de données de la recherche



Virgile Jarrige - 11/2020

Cas d'utilisation – proposition de structure



Dataverse = conteneur

Jeu de données = un ou plusieurs fichiers

Gestion des droits

